

Cloud Services Brokerage: A Survey and Research Roadmap

Adam Barker, Blesson Varghese and Long Thai

School of Computer Science

University of St Andrews, St Andrews, UK.

Email: {adam.barker, varghese, ltt2 }@st-andrews.ac.uk

Abstract—A Cloud Services Brokerage (CSB) acts as an intermediary between cloud service providers (e.g., Amazon and Google) and cloud service end users, providing a number of value adding services. CSBs as a research topic are in there infancy. The goal of this paper is to provide a concise survey of existing CSB technologies in a variety of areas and highlight a roadmap, which details five future opportunities for research.

I. INTRODUCTION

The public cloud computing [1] market is crowded with competitors and choices, each offering their own set of heterogeneous services (e.g., storage, elastic-compute) and configuration options, covering for example: Virtual Machine (VM) instance type (e.g., high-memory, high-CPU etc.), data centre region, availability zone etc. Often it is the case that users deploy applications onto cloud infrastructure on an ad hoc basis, without understanding which providers and instance types can meet user-driven Service Level Objectives (SLOs): individual measurable performance metrics, which cover Quality of Service (QoS) properties such as availability, throughput and response time.

A *Cloud Services Brokerage (CSB)* acts as an intermediary between cloud service providers (e.g., Amazon and Google) and cloud service end users. Brokerages utilise several types of brokers and platforms to enhance service delivery, and ultimately service value [8]. CSBs as a research topic are in there infancy: although there have been numerous research projects which have attempted to develop individual components for a CSB, there have (to date) been no successful integrated frameworks which provide an end-to-end solution. The goal of this paper is to provide a concise survey of existing CSB technologies in a variety of areas and highlight a roadmap, which details five key opportunities for research.

II. SURVEY

Our survey of existing CSB technologies has been broken down into four key categories. Firstly *CSBs for performance*, addresses issues of cloud performance comparison and prediction. *CSBs for application migration*, discusses decision support systems that help guide a developer when making a decision about whether to migrate to the cloud. *Theoretical models for CSBs*, discusses purely theoretical and mathematical techniques. Finally *data for CSBs* presents an overview of providers that supply raw data or metrics, which can be used as input to drive a CSB.

A. CSBs for Performance

STRATOS [20] is a cloud broker service, which focuses on solving a Resource Acquisition Decision (RAD) problem, involving the selection of n resources from m cloud providers. *STRATOS* requires two sets of inputs in order to solve the RAD problem. The first is the application topology to be deployed on the cloud, this is specified through a Topology Descriptor File (TDF). The second is a set of developer objectives: measurable constraints, requirements and preferences. The Service Measurement Index (SMI) [2] is an approach to facilitate the comparison of cloud services. SMI consists of seven categories relating to accountability, agility, assurance, financial, performance, security, privacy and usability, with each category being further refined to a set of attributes, expressed as a set of Key Performance Indicators (KPIs). *STRATOS* utilises KMIs in order express a developer's objectives. *STRATOS* then solves the RAD problem at runtime through a multi-criteria optimisation problem utilising the desired configuration (defined through a TDF), a set of objectives (defined through KMIs) and data from third-party services such as CloudHarmony [4].

OPTIMIS [6] is a multi-cloud toolkit aimed at infrastructure providers, service providers and application developers. It provides a suite of functionalities aimed at optimising IaaS clouds, covering the full lifecycle from service construction, cloud deployment and operation. *OPTIMIS* focuses on key non-functional concerns, regulatory and legislative constraints. A broker component is contained within the *OPTIMIS* toolkit, which allows alternative cloud configurations to be directly compared in terms of business efficiency.

CloudCmp [15] focuses on measuring the performance of a range of cloud services including: elastic compute clusters, persistent storage, intra-cloud networking and wide-area networking. Each broad service area is evaluated through a series of individual benchmarks and metrics. *Computation metrics* are evaluated through Java-based benchmarks from SPECjvm2008, covering the completion time of the benchmark, cost-per-benchmark, and scaling latency: time to start new VMs. *Storage metrics* address the response time, throughput, time to consistency and cost per operation. Finally *networking metrics* are evaluated through standard tools such as iperf and ping.

Conductor [25] is a framework which focuses on optimis-

ing the execution of a MapReduce application on the cloud by utilising resources based on a user's goals, such as minimising cost or completion time. It also supports dynamic monitoring and utilisation in order to handle prediction error or system failure on runtime.

Matrix [3] is a recent research project that is able to predict the performance of a given application on different hardware types based on the behaviour similarity between an application and previously sampled ones. However, it focused on either predicting the performance or selecting the most cost effective instances instead of providing options based on a user's requirements.

CloudBench [23] is a benchmarking methodology and prototype framework in which a user provides a set of four weights that indicate how important each of the following groups: memory, processor, computation and storage are to the application that needs to be executed on the cloud. The weights along with cloud benchmarking data are used to generate a ranking of VMs that can maximise performance of the application.

B. CSBs for Application Migration

CloudProphet [16] is a tool focused on application migration to the cloud. CloudProphet traces (records) the workload of the application when running locally, and replays the same workload in the target cloud(s) for prediction. Once the workload is replayed in multiple target clouds a decision can be made about which cloud to migrate the application; typically this can be a costly and time consuming process.

The Cloud Adoption Toolkit [13] addresses a number of challenges that decision makers face when assessing the feasibility of the adoption of cloud computing in their organisations. The most mature component focuses on cost prediction, and allows an organisation to model their application requirements over time and predict migration and future costs across multiple cloud providers.

RightScale [21] offers a number of services, which allows users to deploy and manage applications across multiple cloud providers. Basic brokering mechanisms are provided through an alert-action mechanism. VMs managed by RightScale have a number of pre-defined hooks, which send data back to the RightScale console, this allows an alert to trigger an associated action, e.g. a scaling policy [10].

C. Theoretical Models for CSBs

SMICloud [7] is a decision support tool, which allows users to evaluate and rank clouds based on a user's Quality of Service (QoS) requirements. SMICloud is based upon the Service Measurement Index (SMI) [2] proposed by the Cloud Service Measurement Index Consortium, also used by STRATOS. SMICloud takes into account functional and non-functional requirements and covers: service response time, sustainability, data centre performance per energy (covering carbon emissions), suitability, accuracy,

transparency, interoperability, availability, reliability, stability, cost, adaptability, elasticity, usability, throughput and scalability. An Analytical Hierarchical Process (AHP) based ranking mechanism is proposed as a way of evaluating user defined QoS requirements against cloud providers.

Authors of [22] *mathematically model* the cloud service selection problem and present a multi-criteria methodology, which focuses on instance selection for performance.

D. Data for CSBs

CloudHarmony [4] is a commercial product, which provides four core services so that users can compare a large set of commercial cloud providers. *CloudSquare* allows users to directly compare the features of cloud providers, e.g., providers offering compute at 100% availability in Europe. *CloudScores* offers a set of benchmarking data for popular cloud providers; currently SPECint, SPECfp, Geekbench, Geekbench Multicore, Unixbench and Unixbench Multicore data are available. *CloudMatch* provides realtime metrics for cloud providers, e.g., networking latency for all providers, upload speed etc. Finally *CloudReports* offers many of the previous services packaged into either quarterly or monthly reports. CloudHarmony have a number of Web service (both REST and SOAP) APIs available, allowing application developers to utilise these metrics for cloud brokers.

A recent *EPSRC project* [9] discussed a set of "fair" benchmarks for cloud computing systems. The report provides a comprehensive survey of existing benchmarking techniques and presents the results of performance benchmarking covering: Memory IO, CPU, Disk IO, Application and Network for Amazon, Rackspace, IBM and a private cloud installation of OpenStack. The research was motivated towards performance-based pricing schemes, e.g., paying more for better, or more reliable performance.

The *Yahoo! Cloud Serving Benchmark (YCSB)* [5] is an open-source framework for evaluating the retrieval and maintenance capabilities of NoSQL database systems.

III. RESEARCH ROADMAP

This Section discusses a set of research opportunities in the form of a roadmap for cloud brokerage. Each opportunity forms an initial step in our idealised broker application. We are proposing further research in the following areas: capturing user-driven requirements, application specific brokering, lightweight cloud benchmarking, dynamic cloud management and a unified Broker as a Service (BaaS) framework.

A. Capturing High-Level User Requirements

Most of the current research efforts on building cloud brokerage systems has focused on a user providing a low-level specification of the hardware required, e.g., two instances with 4Gb of memory, with 4 vCPUs. It is very difficult for users to translate these low-level hardware specifications into

the overall desired performance required by an application. Research needs to address how users specify high-level Service Level Objectives (SLOs), which specify a desirable goal for a deployment, examples of goals include: end-to-end latency, throughput, program execution time, high availability etc. [19] provides an interesting discussion of “good” performance metrics for cloud services brokers.

Research could take the form of a high level policy language and associated toolchain, which translates user defined goals (e.g., maximum end-to-end latency) into a suitable deployment consisting of a set of instance types and geographical regions across multiple cloud providers. Very recent research has also hinted at this problem under the name Software Defined Service Level Agreements (SLAs) [14].

B. Application Specific Brokering

Each user wanting to deploy an application on the cloud has a different (and potentially unique) set of requirements. For example, while a developer of a Hadoop application requires that a certain number of tasks must be completed in a given time frame, an owner of a web server wants her application to be able to respond to any request within a certain amount of time, regardless of the current workload. Similarly, an organisation with a high budget is able to pay a lot of money to minimise its application’s execution runtime, while a small start-up might prefer a cheaper option, which takes longer to complete. As a result, it is necessary to have a cloud broker service, which takes into account an application’s behaviour in order to suggest the best options based on a user’s requirements. As all applications behave differently, the decision made by a broker regarding one application cannot necessarily be applied to another.

Developing a cloud broker which makes decisions based on application specific characteristics is challenging as it requires not only complete knowledge of both the application’s behaviour and cloud resource performance, but also requires a method, which can find the connection between these two data sets in order to make an accurate prediction. We advocate further research upon application modelling and workload classification, capture and prediction. Furthermore, instead of just maximising performance or minimising the execution cost of an application, a broker must be able to enforce a user-defined QoS property based on the application model and predicted workload. CloudProphet [16] and Matrix [3] have taken some first steps towards this goal.

C. Lightweight Cloud Benchmarking

There are a number of efforts to benchmark cloud resources (for example, [23], [17], [11]). Either standard tools or custom-built tools are employed to generate numerous metrics that reveal a number of attributes of cloud resources. Most benchmarking methods are however heavy weight (time consuming) and cannot be used in real-time. Therefore

the data generated has to be used as historic data for predicting the performance of the resource. To be able to make accurate predictions it would be valuable to have historical benchmarking data, which can be used in conjunction with live monitoring and benchmarks.

We propose, that a public (potentially crowdsourced) benchmarking repository be made available for the cloud community, which captures metrics across a range of cloud providers and instance types. By providing applications with historical data across different cloud providers it is more likely that applications can accurately predict and schedule applications given user defined Service Level Objectives (SLOs).

In addition we propose lightweight benchmarking methods to be developed, such that benchmarks can be deployed and executed in near realtime; this is usually in contrast with most benchmarking methods, which as discussed above are time consuming processes. Cloud WorkBench [12] is a first step towards achieving this objective and builds upon the Infrastructure-as-Code concept.

D. Dynamic Cloud Management

During the execution of an application, the broker service must be able to dynamically reschedule an application on the cloud resources to ensure that the high-level requirements of a user are always met, even when the application is still running. For example, assume the user wants to ensure that the end-to-end latency of an application is less than 100 milliseconds, and the network latencies during the execution at the given provider(s) are increasing. The supplied latency user requirement will not be satisfied in the long-term and the broker needs to dynamically monitor and manage the deployment; this might involve: migrating the application onto different resource types, different providers or regions, adding or removing VMs etc. While significant efforts have been made in this direction, for example [24], [18], they still need to be integrated with CSB services in a meaningful way.

E. Broker-as-a-Service (BaaS)

As a future direction, we are proposing a publicly available cloud Broker-as-a-Service (BaaS) framework, which can provide the following across multi-cloud environments: (i) capture high-level user requirements and translate these requirements into concrete resources (instance types, regions etc.), discussed in Section III-A; (ii) provide functionalities for application specific brokering, discussed in Section III-B; (iii) facilitate lightweight benchmarking and utilise benchmarking data through a historical repository, discussed in Section III-C; (iv) and allow dynamic reconfiguration of resources, discussed in Section III-D. The BaaS framework can be made available through a publicly accessible Web-based (e.g., REST) API that can automatically provide a suite of brokering services for applications.

IV. CONCLUSION

This paper has reviewed the current state-of-the-art in Cloud Services Brokerage (CSBs), an important emerging topic in the cloud computing research community. After reviewing the state-of-the-art this paper then presented a research roadmap through an idealised broker, which requires new research contributions in five key areas: capturing and translating high-level user requirements, application specific brokering, new lightweight benchmarking techniques, dynamic cloud management, all of which should be integrated into a unified Broker as a Service (BaaS) framework.

REFERENCES

- [1] A. Barker, B. Varghese, J. S. Ward, and I. Sommerville. Academic cloud computing research: Five pitfalls and five opportunities. In *6th USENIX Workshop on Hot Topics in Cloud Computing, HotCloud '14*. USENIX Association, 2014.
- [2] Carnegie Mellon University. Service measurement index version 1.0. Technical report, 2011.
- [3] R. C. Chiang, J. Hwang, H. H. Huang, and T. Wood. Matrix: Achieving predictable virtual machine performance in the clouds. In *11th International Conference on Autonomic Computing (ICAC 14)*, pages 45–56, Philadelphia, PA, June 2014. USENIX Association.
- [4] CloudHarmony. Cloudharmony. <http://cloudharmony.com>.
- [5] B. F. Cooper, A. Silberstein, E. Tam, R. Ramakrishnan, and R. Sears. Benchmarking cloud serving systems with ycsb. In *Proceedings of the 1st ACM Symposium on Cloud Computing, SoCC '10*, pages 143–154, New York, NY, USA, 2010. ACM.
- [6] A. J. Ferrer, F. Hernández, J. Tordsson, E. Elmroth, A. Ali-Eldin, C. Zsigri, R. Sirvent, J. Guitart, R. M. Badia, K. Djemame, W. Ziegler, T. Dimitrakos, S. K. Nair, G. Kousiouris, K. Konstanteli, T. Varvarigou, B. Hudzia, A. Kipp, S. Wesner, M. Corrales, N. Forgó, T. Sharif, and C. Sheridan. Optimis: A holistic approach to cloud service provisioning. *Future Generation Computer Systems*, 28(1):66 – 77, 2012.
- [7] S. K. Garg, S. Versteeg, and R. Buyya. A framework for ranking of cloud computing services. *Future Gener. Comput. Syst.*, 29(4):1012–1023, June 2013.
- [8] Gartner. Garner says cloud consumers need brokerages to unlock the potential of cloud services. <http://www.gartner.com/newsroom/id/1064712>.
- [9] L. Gillam, B. Li, O. John, A. P. Tomar, et al. Fair benchmarking for cloud computing systems. *Journal of Cloud Computing: Advances, Systems and Applications*, 2(1):6, 2013.
- [10] N. Grozev and R. Buyya. Inter-cloud architectures and application brokering: taxonomy and survey. *Software: Practice and Experience*, 44(3):369–390, 2014.
- [11] A. Iosup, S. Ostermann, N. Yigitbasi, R. Prodan, T. Fahringer, and D. Epema. Performance analysis of cloud computing services for many-tasks scientific computing. *IEEE Trans. Parallel Distrib. Syst.*, 22(6):931–945, June 2011.
- [12] J. C. Joel Scheuner, Philipp Leitner and H. Gall. Cloud workbench: Infrastructure-as-code based cloud benchmarking. In *Proceedings on IEEE CloudCom 2014*, 2014.
- [13] A. Khajeh-Hosseini, D. Greenwood, J. W. Smith, and I. Sommerville. The cloud adoption toolkit: Supporting cloud adoption decisions in the enterprise. *Softw. Pract. Exper.*, 42(4):447–465, Apr. 2012.
- [14] J. Lango. Toward software-defined slas. *Commun. ACM*, 57(1):54–60, Jan. 2014.
- [15] A. Li, X. Yang, S. Kandula, and M. Zhang. Cloudcmp: Comparing public cloud providers. In *Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement, IMC '10*, pages 1–14, New York, NY, USA, 2010. ACM.
- [16] A. Li, X. Zong, S. Kandula, X. Yang, and M. Zhang. Cloudprophet: Towards application performance prediction in cloud. *SIGCOMM Comput. Commun. Rev.*, 41(4):426–427, Aug. 2011.
- [17] C. Luo, J. Zhan, Z. Jia, L. Wang, G. Lu, L. Zhang, C.-Z. Xu, and N. Sun. Cloudrank-d: benchmarking and ranking cloud computing systems for data processing applications. *Frontiers of Computer Science*, 6(4):347–362, 2012.
- [18] S. Meng and L. Liu. Enhanced monitoring-as-a-service for effective cloud management. *IEEE Transactions on Computers*, 62(9):1705–1720, 2013.
- [19] J. O’Loughlin and L. Gillam. Good performance metrics for cloud service brokers. In *The Fifth International Conference on Cloud Computing, GRIDs, and Virtualization*, pages 64–69, 2014.
- [20] P. Pawluk, B. Simmons, M. Smit, M. Litoiu, and S. Mankovski. Introducing stratos: A cloud broker service. In *Cloud Computing (CLOUD), 2012 IEEE 5th International Conference on*, pages 891–898, June 2012.
- [21] RightScale. Rightscale. <http://www.rightscale.com/>.
- [22] Z. ur Rehman, F. Hussain, and O. Hussain. Towards multi-criteria cloud service selection. In *Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS), 2011 Fifth International Conference on*, pages 44–48, June 2011.
- [23] B. Varghese, O. Akgun, I. Miguel, L. Thai, and A. Barker. Cloud Benchmarking for Performance. In *Proceedings of the 6th IEEE International Conference on Cloud Computing Technology and Science (CloudCom 2014)*, pages 535–540, 2014.
- [24] J. Ward and A. Barker. Observing the clouds: a survey and taxonomy of cloud monitoring. *Journal of Cloud Computing*, 3(1), 2014.
- [25] A. Wieder, P. Bhatotia, A. Post, and R. Rodrigues. Orchestrating the deployment of computations in the cloud with conductor. In *Presented as part of the 9th USENIX Symposium on Networked Systems Design and Implementation (NSDI 12)*, pages 367–381, San Jose, CA, 2012. USENIX.